

Transcriptional analysis of in vivo *Plasmodium yoelii* liver stage gene expression[☆]

John B. Sacci Jr.^{a,*}, Jose M.C. Ribeiro^b, Fengying Huang^c, Uzma Alam^a,
Joshua A. Russell^c, Peter L. Blair^c, Adam Witney^{c,d}, Daniel J. Carucci^{c,1},
Abdu F. Azad^a, Joao C. Aguiar^{c,e}

^a Department of Microbiology and Immunology, University of Maryland School of Medicine, 655 W. Baltimore Street, Baltimore, MD 21201, USA

^b Laboratory of Malaria and Vector Research, National Institutes of Health, Bldg. 4 1B2-12, Bethesda, MD 20892, USA

^c Malaria Program, Naval Medical Research Center, Silver Spring, MD 20910, USA

^d Department of Cellular and Molecular Medicine, St. George's Hospital Medical School, London SW17 ORE, UK

^e Pan American Health Organization, USA

Received 28 December 2004; received in revised form 7 March 2005; accepted 31 March 2005

Available online 26 April 2005

Abstract

The transcriptional repertoire of the in vivo liver stage of *Plasmodium* has remained largely unidentified and seemingly not amenable to traditional molecular analysis because of the small number of parasites and large number of uninfected hepatocytes. We have overcome this obstruction by utilizing laser capture microdissection to provide a high quality source of parasite mRNA for the construction of a liver stage cDNA library. Sequencing and annotation of this library demonstrated expression of 623 different *Plasmodium yoelii* genes during development in the hepatocyte. Of these genes, 25% appear to be unique to the liver stage. This is the first comprehensive analysis of in vivo gene expression undertaken for the liver stage of *P. yoelii*, and provides insights into the differential expression of *P. yoelii* genes during this critical stage of development.

© 2005 Elsevier B.V. All rights reserved.

Keywords: Liver stage; *Plasmodium yoelii*; In vivo; cDNA library

1. Introduction

Parasites belonging to the genus *Plasmodium* have a complex lifecycle that requires the stage specific expression of

different proteins to complete development in both the insect vector and vertebrate host. It has been intuitively and empirically understood that the parasites express different sets of proteins during development and that these proteins can be shared or unique to specific stages. There are numerous examples of the stage specific expression of parasite genes at the sporozoite and blood stage [1–5], but experimental evidence identifying the breadth of this regulated gene expression has, until recently, been lacking. However, recent work has shown that cDNA libraries can be constructed to generate expressed sequence tags (ESTs) or expression profiling can be done by high-density oligonucleotide array analysis. These methodologies can provide significant information about gene expression [6,7], with the EST library results having been used to define the transcriptome of the

[☆] The opinions expressed are those of the authors and do not reflect the official policy of the Department of the Navy, Department of Defense, or the U.S. government. The experiments reported here were carried out according to the principles set forth in the "Guide for the Care and Use of Laboratory Animals", Institute of Laboratory Animals Resources, National Research Council, National Academy Press, 1996.

* Corresponding author. Tel.: +1 410 706 4071; fax: +1 410 706 0282.

E-mail address: jsacc001@umaryland.edu (J.B. Sacci Jr.).

¹ Present address: Grand Challenges in Global Health Initiative, Foundation for the National Institutes of Health, 45 Center Drive (3AN-44), Bethesda, MD 20892-460, USA.

Report Documentation Page			<i>Form Approved OMB No. 0704-0188</i>					
<p>Public reporting burden for the collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to a penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number.</p>								
1. REPORT DATE 2005	2. REPORT TYPE N/A	3. DATES COVERED -						
4. TITLE AND SUBTITLE Transcriptional analysis of in vivo Plasmodium yoelii liver stage gene expression			5a. CONTRACT NUMBER					
			5b. GRANT NUMBER					
			5c. PROGRAM ELEMENT NUMBER					
6. AUTHOR(S)			5d. PROJECT NUMBER					
			5e. TASK NUMBER					
			5f. WORK UNIT NUMBER					
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) Department of Microbiology and Immunology, University of Maryland School of Medicine, 655 W. Baltimore Street, Baltimore, MD 21201			8. PERFORMING ORGANIZATION REPORT NUMBER					
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES) Naval Medical Research Center 503 Robert Grant Avenue Silver Spring, MD 20910-7500			10. SPONSOR/MONITOR'S ACRONYM(S)					
			11. SPONSOR/MONITOR'S REPORT NUMBER(S)					
12. DISTRIBUTION/AVAILABILITY STATEMENT Approved for public release, distribution unlimited								
13. SUPPLEMENTARY NOTES The original document contains color images.								
14. ABSTRACT								
15. SUBJECT TERMS								
16. SECURITY CLASSIFICATION OF: <table border="1" style="margin-left: auto; margin-right: auto;"> <tr> <td style="text-align: center;">a. REPORT unclassified</td> <td style="text-align: center;">b. ABSTRACT unclassified</td> <td style="text-align: center;">c. THIS PAGE unclassified</td> </tr> </table>			a. REPORT unclassified	b. ABSTRACT unclassified	c. THIS PAGE unclassified	17. LIMITATION OF ABSTRACT SAR	18. NUMBER OF PAGES 7	19a. NAME OF RESPONSIBLE PERSON
a. REPORT unclassified	b. ABSTRACT unclassified	c. THIS PAGE unclassified						

parasites during their development [8]. The libraries have been quite useful, but it has been the completed shotgun sequencing of several *Plasmodium* genomes [9,10], which has truly allowed the effective utilization of these sequence libraries. The resulting database has become available and can be effectively used for gene discovery and the identification of gene expression from the sporozoite and blood stage.

The utilization of the *Plasmodium yoelii* rodent model for malaria has facilitated the identification of stage specific gene expression and allowed the identification of many parasite genes [11–14]. A significant amount of data now exists to support the studies of gene expression during the sporozoite and blood stages of the parasites' development. Unfortunately, gene expression analysis during *in vivo* liver stage development has not been amenable to the types of molecular investigations that would serve to identify differentially expressed genes [15]. One of the major impediments in identifying parasite gene expression at the liver stage has remained the large number of uninfected hepatocytes relative to the number of infected hepatocytes in the liver after sporozoite inoculation. Isolation of RNA from an infected liver results in an abundance of host RNA and a minor, though detectable [16] amount of parasite message. Current work by one group, to get around this problem, has utilized an *in vitro* culture method to transform sporozoites into early exo-erythrocytic forms (EEFs) and subsequently make an EST library from isolated parasite RNA [17]. We chose an approach that allowed us to look at the *in vivo* expression of parasite mRNA during liver stage development. Recent innovations in microdissection have dramatically improved our abilities to separate schizonts from hepatocytes. Work from our laboratory demonstrated that laser capture microdissection (LCM) could be used to recover liver stage parasites from *P. yoelii* infected mouse liver and that the recovered schizonts were virtually free of contaminating host tissue [18]. These results clearly presented an opportunity to construct a cDNA library, from isolated schizonts that would be almost exclusively derived from parasite mRNA and thus avoid the contamination of the library with host derived sequences.

We now report the generation, sequencing and annotation of a *P. yoelii* liver stage EST library that has supplied significant information about the *in vivo* expression of *P. yoelii* genes during this critical stage of the parasite's lifecycle. This is the first analysis of gene expression during *in vivo* liver stage development and as such has allowed the initial identification of both constitutively and developmentally regulated gene transcripts as they appear in the infected hepatocyte. This work complements and extends the previous studies done to identify the transcriptome and proteome expressed during other stages of the parasites' development. Taken together, the analysis of *P. yoelii* liver stage gene expression offers a unique view of the biology of plasmodial parasites and the possibility of identifying regulated gene function throughout the parasite's lifecycle.

2. Materials and methods

2.1. Parasite preparation

A BALB/c mouse was infected with 5×10^6 of *P. yoelii* 17 × NL sporozoites. The Infected liver was harvested 40 h later and processed for cryosectioning as previously reported [18].

2.2. Laser capture microdissection and RNA extraction

Laser capture microdissection of liver stage schizonts was accomplished using a Pixcell laser capture microscope equipped with an infrared diode laser (Arcturus Engineering, Santa Clara, CA) as previously reported [18].

Briefly, tissue sections on glass slides, were fixed in 70% ethanol, stained with hematoxylin and eosin, dehydrated and air-dried. RNase inhibitor (SUPERase In™, Ambion, Austin, TX) was included in all aqueous staining and washing steps to limit the possibility of RNA degradation. Each dehydrated tissue section was placed on the microscope stage and a transfer cap placed on the tissue. Individual schizonts were captured onto the thermoplastic membrane, mounted onto the transfer cap, by melting of the membrane via laser activation. Typically, the laser spot size was 7.5 μm , the power 40 mW and the pulse duration 3 m s.

Total RNA, from cohorts of caps containing 1500 laser microdissected schizonts, was isolated using the StrataPrep Total RNA Microprep Kit from Stratagene (LaJolla, California) following the manufacturers' directions.

2.3. cDNA library construction and sequencing

Total RNA, from LCM recovered parasites, was Dnase I treated to remove possible genomic DNA contamination. Total RNA was transcribed into cDNA and amplified using the SMART™ cDNA Synthesis kit (Clontech, Palo Alto, CA). After amplification for 18 cycles, cDNA synthesis was evaluated by running on an agarose gel and then digested with Sfi I and size fractionated on a CHROMA SPIN column. Only fractions showing cDNA larger than 300 bp were pooled and ligated into Sfi I-digested λ TriplEx2 Vector (Clontech, Palo Alto, CA). The cDNA library was packaged using the Gigapack III Gold Packaging extract (Stratagene, La Jolla, CA). Single phage plaques were picked and converted into pTriplEx2 plasmid clones using BM25.8 host cells and a protocol recommended by Clontech. A preliminary evaluation of 96 clones indicated that the cDNA library contained approximately 80% recombinant clones. DNA was isolated from 2628 of clones using a mini-prep kit from Millipore (Bedford, MA) and sequenced using both pTriplEx2-specific forward and reverse primers and the BigDye terminator method on a 3100 Genetic Analyzer (ABI, Foster City, CA).

2.4. Assemblies and database searches

Sequences were trimmed from their vector and primer contaminants using a program written in Visual Basic, as described before [19]. Only those sequences having >99 nucleotides (nt) and less than 10% N (undetermined nucleotide base) were used. Trimmed sequences were clustered by first comparing the set against itself using BlastN [20] without filters (FF options), and clustering the sequences presenting at least 90% identity over a 100 nucleotide (nt) length. These clusters were further assembled with the CAP program [21], which could generate more than one CAP cluster per Blast cluster. The assembled sequence data was further compared to several other databases, using Blastx, BlastN or rpsBlast, as follows. The non-redundant protein database set of the National Center for Biotechnology Information (NCBI), a group of human and *Drosophila* protein sequences annotated by the Gene Ontology consortium [22], the Conserved Domains Database of the NCBI (<http://www.ncbi.nlm.nih.gov/Sitemap/index.html#CDD>) containing the PFAM [23], SMART [24], and COGs [25] alignments of protein motifs, three subsets of the NCBI nucleotide databases containing all mitochondrial, rRNA and mouse EST sequences, all *Plasmodium falciparum* protein sequences deposited on NCBI, the nucleotide sequences for the genome and CDS of *P. yoelii* found for public download at ftp://ftp.tigr.org/pub/data/Eukaryotic_Projects/p_yoelii/, and the stage specific cDNA sequences from the blood- and mosquito salivary gland-stages of *P. yoelii* [8]. The results of these Blast matches were directed to a hyperlinked Excel spreadsheet that facilitated the task of manually annotating the database. In order to compare gene expression of *P. yoelii* in the three stage specific libraries, the blood and salivary gland databases were submitted to the same procedures (vector trimming and clustering) as done for the liver stage library, and the resulting contigs of each library blasted against the others with a cutoff of 1E-15. Blast searches were done locally from executables obtained at the NCBI FTP site (<ftp://ftp.ncbi.nih.gov/blast/executables/>). Using similar methodologies to those described above, an analysis was done to identify *P. falciparum* homologs in the *P. yoelii* EST library. As in the initial Blast analysis, a cutoff of 1E-15 was used.

2.5. RT-PCR analysis

RNA for use in RT-PCR analysis, was isolated from *P. yoelii* sporozoites, infected liver and infected blood as previously described [16]. First-strand cDNA was generated from total RNA using the SuperScript First-Strand Synthesis System for RT-PCR kit (Life Technologies, Gaithersburg, MD). The synthesis of cDNA was performed by priming RNA, isolated from the different parasite stages with either random hexamers or oligo-dT, and then incubation with reverse transcriptase (RT+). As a control for the presence of genomic DNA, reactions were done omitting the reverse transcriptase

(RT-). Amplification of specific gene sequences was accomplished by PCR using a hot start Taq DNA polymerase from the HotStarTaq PCR kit (Qiagen, Valencia, CA). One microliter from the cDNA reaction was added to a PCR master mix with one of several oligonucleotide primer pairs (Table S6). The primers were designed and the specific conditions used for the PCR reactions were determined by using MacVector sequence analysis software (Accelrys, San Diego, CA). PCR products were electrophoresed on a 1% agarose gel, stained with ethidium bromide and the DNA bands visualized with an AlphaImager™ (Alpha Innotech Corporation, San Leandro, CA). RT-PCR products were cloned into the TA cloning vector and sequenced to verify that they matched to the targeted gene.

3. Results

3.1. Isolation of *P. yoelii* liver stage schizonts, RNA extraction and cDNA library construction

P. yoelii schizonts were captured by LCM from mouse liver tissue 40 h after sporozoite infection as previously reported [18]. Parasites were not fully mature but could be easily visualized by hematoxylin and eosin staining. Total RNA was extracted from approximately 1500 LCM schizonts and a cDNA library constructed from 1 µg of total RNA.

Isolated RNA was reverse transcribed and amplified using the SMART™ kit. A portion of the amplified material was assessed by agarose gel electrophoresis (Fig. 1), demonstrating the production of significant product above 300 bp that was then size fractionated for ligation into the ITriplex2 vector.

3.2. EST library analysis

A total of 2628 clones were sequenced with 1661 clones containing good quality readable recombinant inserts. The average length of each insert was 413 bp. All of the ESTs have been deposited with GenBank (accession nos. CK573044–CK574949).

The 1661 EST sequences were grouped into 1345 clusters, based upon ESTs having overlapping sequence. These clusters, containing one or more overlapping sequences, were then assembled into contigs. Only 1% of the EST clusters matched mouse sequence, while 91% matched the *P. yoelii* genome and 72.4% matched the predicted CDS of *P. yoelii* (Table 1). From the contigs, 623 non-redundant genes were identified. The difference between the number of matches for the genome and predicted genes may be due to ESTs from UTRs or, less likely, ESTs from unidentified genes. The majority of the sequences that had no match to the database, were of relatively low complexity and had GC content percentages that were 50% or more. The AT content of *P. yoelii* genes is generally greater than 70%, suggesting that these low complexity sequences were not derived from parasite genes. Curiously, a small number, six sequences, of the ESTs that

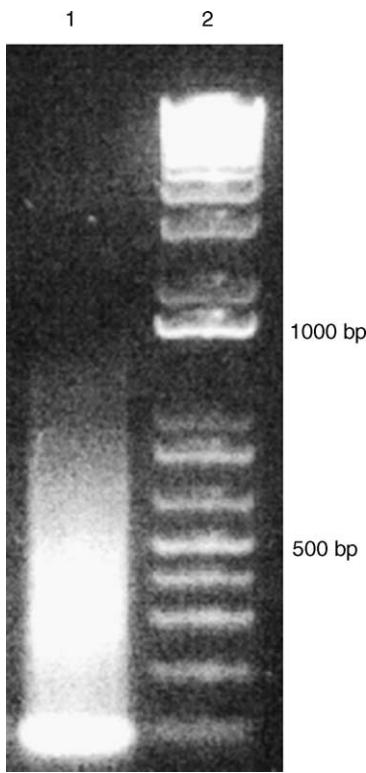


Fig. 1. Agarose gel analysis of linearly amplified cDNA. RNA, isolated from 40 h liver schizonts was reverse transcribed and then amplified using the SMART™ cDNA Synthesis kit (Clontech, Palo Alto, CA). Lane 1, amplified cDNA from LCM isolated *Plasmodium yoelii* schizonts. Lane 2, molecular weight markers.

did not have significant homology to *P. yoelii* genomic sequence did have high homology to *P. falciparum* genomic sequences. Internal primers were constructed from three of the ESTs (Py EST 248, Py EST 516 and Py EST 892) and used to successfully amplify mRNA from liver stage parasites. These sequences appear to represent genes that were missed in the initial annotation of the *P. yoelii* genome shotgun sequencing project [9].

To define the differential expression of *P. yoelii* genes, the EST sequences obtained from mosquito salivary gland, liver, and blood stage cDNA libraries were Blast analyzed against the predicted CDS of *P. yoelii* with a BlastN cutoff of 1E – 12, thus identifying genes expressed in all three libraries, or in

any other combination. The complete tables listing the stage distribution of the ESTs, and their corresponding genes, are in Tables S4–S7 found in supporting information online. These tables have been hyperlinked to appropriate databases to allow efficient utilization of the EST library to identify genes or proteins.

This approach of comparing of identified genes was preferred rather than comparing the EST sequences to each other, because multiple ESTs could map to different regions of the same gene, but show no match to each other. This could potentially bias the analysis and make stage specific comparisons misleading. Our analysis also eliminated problems that could derive from differences in library construction, such as biasing or not for 5' regions of mRNA. The distribution of the identified genes among the different developmental stages was consistent with the time of development at which the liver parasites were harvested. A very small percentage (1.9%) of the total number of genes were found to match expressed genes previously found uniquely in the sporozoite library, while a higher percentage (16%) had significant matches to both sporozoite and blood stage sequences. As might be expected, based upon the time in development when the parasites were harvested, a large percentage of the genes (56.6%) had matches to blood stage genes identified by the blood stage EST library (Table 2). The most significant finding, however, was that 25% of the genes identified by the liver stage ESTs, did not match either sporozoite or blood stage expressed genes. This suggested that a considerable fraction of these ESTs, from our library, were liver stage specific and represented message for proteins that were uniquely expressed during this stage of development. This finding that a significant fraction of the library represents potential stage specific expression is similar to the results seen in a proteomic analysis of multiple *P. falciparum* stages [26]. Those results identified between 20 and 33% of the blood stage proteins as being specific to trophozoites, merozoites or gametocytes.

3.3. RT-PCR analysis

RT-PCR analysis was done to confirm that the expression patterns identified by comparing the EST libraries, from the

Table 2

Genes identified from the *P. yoelii* liver stage library demonstrating significant matches to genes identified by Blast analysis of *P. yoelii* sporozoite or blood stage EST libraries (based on BlastN cutoff of 1e – 12)

Database	Number of ESTs (%)	Number of clusters (%)	AT (%)	Number of matches	Total (%)
Mouse	19 (1.1)	16 (1.1)	55.6	12 ^a	1.9 ^b
rRNA	41 (2.4)	12 (0.9)	60.9	353	56.7
Py Genome	1508 (90.8)	1226 (91.1)	70.7	100	16
Py CDS	1269 (76.3)	975 (72.4)	70.4	158	25.3
Pf proteins ^a	6 (0.3)	6 (0.4)	78.2		
No match ^b	87 (5)	85 (6)	51.9	623	100

Table 1

Distribution of *P. yoelii* sequences and clusters

Database	Number of ESTs (%)	Number of clusters (%)	AT (%)
Mouse	19 (1.1)	16 (1.1)	55.6
rRNA	41 (2.4)	12 (0.9)	60.9
Py Genome	1508 (90.8)	1226 (91.1)	70.7
Py CDS	1269 (76.3)	975 (72.4)	70.4
Pf proteins ^a	6 (0.3)	6 (0.4)	78.2
No match ^b	87 (5)	85 (6)	51.9

^a ESTs that did not match sequence in the *P. yoelii* genomic database, but did match to the *P. falciparum* proteome (by Blastx), with a cutoff of 1e – 15.

^b ESTs that did not demonstrate a significant match to sequence in any database.

^a Number of non-redundant predicted genes.

^b Percent of non-redundant genes identified in a specific subset in relation to all of the non-redundant genes identified by the liver stage cDNA library.

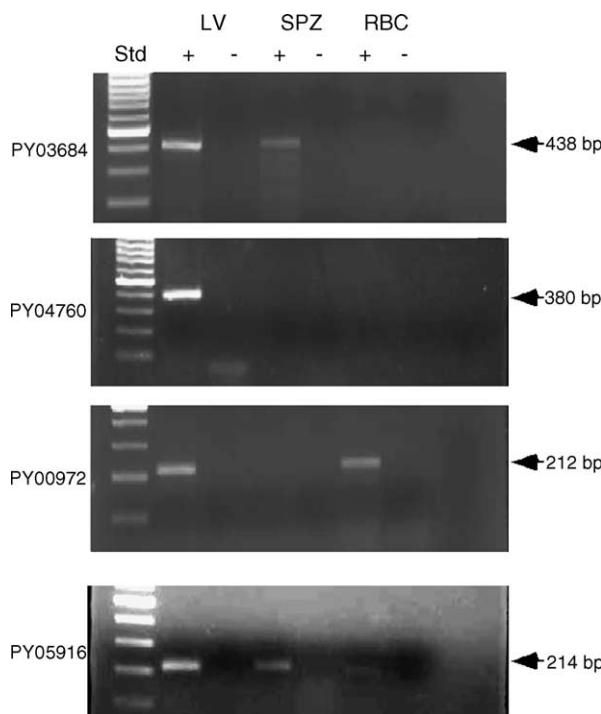


Fig. 2. Representative RT-PCR analysis of *P. yoelii* gene sequences present in the in vivo liver stage EST library. RT-PCR analysis was done using primers for genes identified from the Blast searches as being expressed during sporozoite and liver stages (PY03684), liver stage only (PY04760), liver and blood stages (PY00972) and sporozoite, liver and blood stages (PY05916). Total RNA from infected liver (LV), sporozoites (SPZ) and infected red blood cells (RBC) was subjected to first strand synthesis with (+) or without (−) reverse transcriptase. The cDNA was then used as the template for PCR with the gene specific primers.

different stages of the parasite, were accurate. As it would have been impractical to analyze all of the genes, a small group of genes were selected for amplification from the different expression groups (sporozoite and liver; liver only, liver, blood stage and sporozoite; liver and blood stage). Genes with putative identifications as well as genes that coded for proteins with unknown function were included (Table S9). Additionally, some of the genes were chosen because they contained putative signal sequences, transmembrane domains or represented potential virulence factors. The primer sequences that were designed to amplify the 3' regions of the selected transcripts are detailed in Table S10. Representative results for the RT-PCR analysis are presented in Fig. 2. In all cases, controls were done in which first strand synthesis did not contain reverse transcriptase. These were uniformly negative, demonstrating that the PCR products were the result of amplification from cDNA. Additionally, the PCR products were cloned and sequenced to verify that the products matched the gene being assessed.

Expression patterns in the different stages were, for the most part, consistent with those predicted from the database (Table 3). In several cases, however, genes that were predicted to have expression in specific stages did not produce RT-PCR products or were RT-PCR positive when there was no predic-

Table 3
RT-PCR analysis of *P. yoelii* predicted gene expression

Gene locus ^a	RT-PCR		
	Sporozoite	Liver stage	Blood stage
SPZ–LV^b			
PY00150	+	+	
PY03269	+	+	
PY04496	+	+	
PY04658	+	+	
LV			
PY01013		+	
PY01200		+	
PY04760		+	
PY05129		+	
PY05811		+	
PY06882	+	+	
LV–BLD			
PY00516		+	
PY00487		+	+
PY00972		+	+
PY03267		+	+
PY04945		+	+
PY07320		+	+
SPZ–LV–BLD			
PY00479	+	+	
PY01412		+	+
PY03207		+	+
PY05916	+	+	+

^a The Institute for Genome Research Locus identifier for predicted *P. yoelii* coding regions.

^b Stage specific expression of putative genes as identified from EST database analysis of sporozoite (SPZ), liver (LV) and blood (BLD).

tion for gene expression. It is possible that these transcripts were of low abundance in the sporozoites or blood stages, requiring higher amounts of cDNA template to produce a positive PCR product and in the case of positives where no product was predicted, it is certainly possible that expression was not represented in the EST libraries. Of interest was the liver stage specific expression of the six genes chosen from the database for RT-PCR validation. These genes were predicted to be expressed in only the liver stage and it was encouraging to find that 5 of 6 produced a positive RT-PCR product from liver stage material.

3.4. Categorization of putative function

The non-redundant protein database was searched by Blastx analysis, and possible functions for the *P. yoelii* liver stage clusters were assigned (Fig. 3). As described in previous reports of functional classifications for other stages of the parasite's lifecycle [6,7], a majority of the ESTs (58%) represented proteins with an unknown function. Genes involved in protein synthesis and nuclear regulation, while representing a relatively small percentage of the total, were the next most common groups. This was not surprising, as development is very rapid and the processes that support the rapid growth of the parasite would be significantly represented. These findings were in contrast to the results obtained from a *P. yoelii*

Functional Classification of *P. yoelii* Liver Stage ESTs

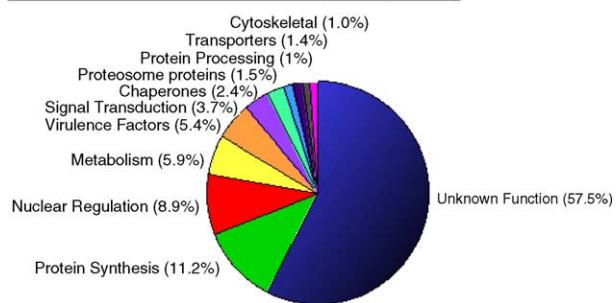


Fig. 3. Functional profile of *P. yoelii* genes identified by liver stage EST library. The results of a Blast analysis of the *P. yoelii* liver stage EST library identified 632 genes that were expressed during development in the hepatocyte. Further Blastx evaluation assigned a putative biological function to the matches and the genes were then clustered and are presented as a percentage of the total. The specific ascribed function for an expressed gene can be found in the supplemental tables.

sporozoite library, in which it was found that ribosomal proteins and other proteins involved in protein synthesis were underrepresented [8]. This is not unexpected, as sporozoites are developmentally mature and would not have nearly the same level of transcriptional and translational activity as parasites in a developing schizont. Chaperones and heat shock proteins made up a relatively small portion of the ESTs.

3.5. Identification of orthologous *P. falciparum* genes

Blastx analysis, of the EST library, was done to identify potential orthologous *P. falciparum* genes. Using a cut off value of $1e - 15$, 438 EST clusters were found to have significant matches to *P. falciparum* transcripts. A spreadsheet in the supporting online material (Table S8) identifies the *P. yoelii* gene locus and the corresponding homologous *P. falciparum* locus. As might be expected, there were a number of genes involved in DNA replication and protein synthesis (ribosomal proteins predominantly) as well as other house-keeping functions that shared a high degree of homology between the two species. As seen in assessing the putative function of the *P. yoelii* genes, almost half of the identified orthologues coded for hypothetical proteins.

4. Discussion

Infected hepatocytes are an important target of a protective immune response during a *P. yoelii* infection [27–29]. This finding makes the identification and characterization of plasmodial antigens expressed in the infected hepatocyte and the immune responses against these antigens crucial for the development of a pre-erythrocytic stage malaria vaccine. Additionally, understanding the biology associated with the development of the parasite in the hepatocyte will not only further our understanding of this part of the lifecycle, but may provide insights into the development of effective chemotherapeutics. Our initial identification of in vivo liver stage gene

expression has demonstrated that this stage of the parasites' lifecycle is amenable to molecular analysis and that the in vivo differential expression of the plasmodial liver stage transcriptome can be assessed.

Work published recently by Wang et al. [17] used an in vitro manipulation of *P. yoelii* sporozoites to produce early exo-erythrocytic stage parasites from which a cDNA library was generated. While there were some similarities between that library and our EST library, there were also substantial differences that were most likely related to the parasite stage from which RNA was extracted. A Blast analysis comparing the two libraries found 98 genes that were common to both libraries and also found in the available databases for *P. yoelii* sporozoites and blood stages. Additionally, ESTs from both libraries identified 21 genes that were not found in any of the *P. yoelii* expression databases (putatively liver stage specific), but were common to our library and the Wang library. As reported by Wang, we also detected expression of message for the *P. yoelii* PfEMP3-like molecules, merozoite surface protein-1 (MSP-1) and ring-infected erythrocyte surface antigen (RESA), but unlike Wang found no ESTs in our library for the circumsporozoite protein (CS) and sporozoite surface protein 2 (SSP2). This is consistent with our previous work, in which we were unable to amplify message for CS from LCM liver stage schizonts [18]. Additionally, we had representation of PyHEP17 message, an antigen that is the *P. yoelii* orthologue of *P. falciparum* exported protein-1 (EXP-1). PyHEP17 is expressed very early during liver stage development and subsequently into the blood stage. We also found some expression of message for heat shock proteins and chaperones, but they represented only a little over 2% of the ESTs. The presence of ESTs that represented a broad range of gene expression and the lack of many sporozoite transcriptional elements suggested that this library was representative of the maturing in vivo liver stage transcriptome.

While the EST library described in this work was made from relatively late stage parasites (~ 40 h) we have recently been able to microdissect parasites from as early as 30 h after infection (data not shown) by using an immunofluorescent staining protocol to visualize the early developing schizonts. The comparison of transcription at this earlier stage with the more mature parasites may provide insights into the changes that occur in gene expression as the parasite goes through the transitional liver stage.

The most interesting and difficult aspect of this analysis has been the large number of genes identified that code for proteins with no known function. These genes are of interest, as they do not appear in any database and represent potentially unique targets. The recent publication of methodologies to rapidly test hundreds of unknown genes [30,31] will likely provide information about these proteins with unknown functions and aid in the selection of drug and vaccine targets. A comprehensive analysis, of these genes, is beyond the scope of this initial work and will require extensive future studies. It is clear that there is a tremendous potential for thorough gene analysis at the liver stage and it should now be possible to

utilize the genomic data that has been made available through genome sequencing projects to accomplish this aim.

Acknowledgments

This work was supported by the NIH (RO1 AI 47445) and by the Naval Medical Research Center and the Military Infectious Disease Research Program work unit # A0262. The authors would like to acknowledge the expert assistance of Calvin Williams in performing some of the RT-PCR assays.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.molbiopara.2005.03.018.

References

[1] Rogers WO, Malik A, Mellouk S, et al. Characterization of *Plasmodium falciparum* sporozoite surface protein 2. Proc Natl Acad Sci USA 1992;89:9176–80.

[2] Fidock DA, Bottius E, Brahim KA, et al. Cloning and characterization of a novel *Plasmodium falciparum* sporozoite surface antigen. STARP Mol Biochem Parasitol 1994;64:219–32.

[3] Dame JB, Williams JL, McCutchan TF, et al. Structure of the gene encoding the immunodominant surface antigen on the sporozoite of the human malaria parasite *Plasmodium falciparum*. Science 1984;225:593–9.

[4] Ling IT, Kaneko O, Narum DL, et al. Characterisation of the rhoph2 gene of *Plasmodium falciparum* and *Plasmodium yoelii*. Mol Biochem Parasitol 2003;127:47–57.

[5] Howard RF, Narum DL, Blackman M, Thurman J. Analysis of the processing of *Plasmodium falciparum* rhoptry-associated protein 1 and localization of Pr86 to schizont rhoptries and p67 to free merozoites. Mol Biochem Parasitol 1998;92:111–22.

[6] Le Roche KG, Zhou Y, Blair PL, et al. Discovery of gene function by expression profiling of the malaria parasite life cycle. Science 2003;301:1503–8.

[7] Bozdech Z, Llinás M, Pulliam BL, Wong ED, Zhu J, DeRisi JL. The transcriptome of the intraerythrocytic developmental cycle of *Plasmodium falciparum*. PloS Biol 2003;1:85–100.

[8] Kappe SH, Gardner MJ, Brown SM, et al. Exploring the transcriptome of the malaria sporozoite stage. Proc Natl Acad Sci USA 2001;98:900–895.

[9] Carlton JM, Angiuoli SV, Suh BB, et al. Genome sequence and comparative analysis of the model rodent malaria parasite *Plasmodium yoelii*. Nature 2002;419:512–9.

[10] Gardner MJ, Hall N, Fung E, et al. Genome sequence of the human malaria parasite *Plasmodium falciparum*. Nature 2002;419:498–511.

[11] Kappe SH, Noe AR, Fraser TS, Blair PL, Adams JH. A family of chimeric erythrocyte binding proteins of malaria parasites. Proc Natl Acad Sci USA 1998;95:1230–5.

[12] Burns JM, Belk CC, Dunn PD. A protective glycosylphosphatidylinositol-anchored membrane protein of *Plasmodium yoelii* trophozoites and merozoites contains two epidermal growth factor-like domains. Infect Immunol 2000;68:6189–95.

[13] Doolan DL, Hedstrom RC, Rogers WO. Identification and characterization of the protective hepatocyte erythrocyte protein 17 kDa gene of *Plasmodium yoelii*, homolog of *Plasmodium falciparum* exported protein 1. J Biol Chem 1996;271:17861–8.

[14] Preiser PR, Khan S, Costa FTM, et al. Stage-specific transcription of distinct repertoires of a multigene family during *Plasmodium* life cycle. Science 2002;295:342–5.

[15] Gruner AC, Brahim K, Eling W, et al. The *Plasmodium falciparum* knob-associated PfEMP3 antigen is also expressed at pre-erythrocytic stages and induces antibodies which inhibit sporozoite invasion. Mol Biochem Parasitol 2001;112:253–61.

[16] Lau AOT, Sacci Jr JB, Azad AF. Detection of *Plasmodium yoelii* liver stage mRNA in BALB/c mice. J Parasitol 2001;87:19–23.

[17] Wang Q, Brown S, Roos DS, Nussenzeig V, Bhanot P. Transcriptome of axenic liver stages of *Plasmodium yoelii*. Mol Biochem Parasitol 2004;137:161–8.

[18] Sacci Jr JB, Aguiar JC, Lau AO, Hoffman SL. Laser capture microdissection and molecular analysis of *Plasmodium yoelii* liver-stage parasites. Mol Biochem Parasitol 2002;119:285–9.

[19] Valenzuela JG, Pham VM, Garfield MK, Francischetti IM, Ribeiro JMC. Toward a description of the sialome of the adult female mosquito *Aedes aegypti*. Insect Biochem Mol Biol 2002;32:1101–22.

[20] Altschul SF, Madden TL, Schaffer AA. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucl Acids Res 1997;25:3389–402.

[21] Huang X. An improved sequence assembly program. Genomics 1992;33:21–31.

[22] Ashburner M, Ball CA, Blake JA, et al. Gene ontology: tool for the unification of biology. The gene ontology consortium. Nat Genet 2000;25:25–9.

[23] Bateman A, Birney E, Durbin R, Eddy SR, Howe KL, Sonnhammer EL. The Pfam protein families database. Nucl Acids Res 2000;28:263–6.

[24] Schultz J, Copley RR, Doerks T, Ponting CP, Bork P. SMART: a web-based tool for the study of genetically mobile domains. Nucl Acids Res 2000;28:231–4.

[25] Natale DA, Galperin MY, Tatusov RL, Koonin EV. Using the COG database to improve gene recognition in complete genomes. Genetica 2000;108:9–17.

[26] Florens L, Washburn MP, Raine JD, et al. A proteomic view of the *Plasmodium falciparum* life cycle. Nature 2002;419:520–6.

[27] Hoffman SL, Isenbarger D, Long GW, et al. Sporozoite vaccine induces genetically restricted T cell elimination of malaria from hepatocytes. Science 1989;244:1078–81.

[28] Hoffman SL, Weiss W, Mellouk S, Sedegah M. Irradiated sporozoite vaccine induces cytotoxic T lymphocytes that recognize malaria antigens on the surface of infected hepatocytes. Immunol Lett 1990;25:33–8.

[29] Weiss WR, Sedegah M, Beaudoin RL, Miller LH, Good MF. CD8+T cells (cytotoxic/suppressors) are required for protection in mice immunized with malaria sporozoites. Proc Natl Acad Sci 1988;85:573–6.

[30] Haddad D, Bilcikova E, Witney AA, et al. Novel antigen identification method for discovery of protective malaria antigens by rapid testing of DNA vaccines encoding exons from the parasite genome. Infect Immunol 2004;72:1594–602.

[31] Aguiar JC, LaBaer J, Blair PL, et al. High-throughput generation of *P. falciparum* functional molecules by recombinational cloning. Genome Res 2004;14:2076–82.